BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 9: e147432
doi: 10.3897/biss.9.147432

OPEN ACCESS

Conference Abstract

Supporting Biodiversity Dataset Preparation: An Introduction to the TaiBIF Open Data Toolkit

Melissa Jean-Yi Liu[‡,§], Jhu-Jyun Jhang[‡], Daphne Hoh[‡], Chun-I Chang[‡], Mao-Ning Tuanmu[‡]

‡ Taiwan Biodiversity Information Facility, Biodiversity Research Centre, Academia Sinica, Taipei, Taiwan
§ Global Biodiversity Information Facility, Taipei, Taiwan

Corresponding author: Mao-Ning Tuanmu (mntuanmu@gate.sinica.edu.tw)

Received: 22 Jan 2025 | Published: 22 Jan 2025

Citation: Liu MJ-Y, Jhang J-J, Hoh D, Chang C-I, Tuanmu M-N (2025) Supporting Biodiversity Dataset Preparation: An Introduction to the TaiBIF Open Data Toolkit. Biodiversity Information Science and Standards 9: e147432. https://doi.org/10.3897/biss.9.147432

Abstract

The Darwin Core standard (DwC; Darwin Core Task Group 2009) and the Global Biodiversity Information Facility (GBIF) new data model provide a flexible set of biodiversity data fields to accommodate various thematic datasets such as taxonomic checklists, sampling events, eDNA metabarcoding, ecological survey data, etc. However, this flexibility can make it challenging for data providers to get started, often leading to frustration when trying to map their original data fields to DwC terms. Additionally, while data cleaning is crucial for enhancing data quality (Chapman 2005), it requires significant expertise and effort, which may hinder the mobilization of high-quality data. To address these common pain points in data mobilization, the Taiwan Biodiversity Information Facility (TaiBIF) developed the TaiBIF Open Data Toolkit (ODT) by integrating various thematic dataset templates, DwC terms, the Nansen Legacy Excel Template Generator, Excel data editing interfaces (e.g, filter, cell editing, AutoFill), GBIF Data Validator, and the OpenRefine. We combined the strengths of each tool to create a straightforward workflow from data sheet generation (Fig. 1, Fig. 2), data validation (Fig. 3), data cleaning to dataset packaging (Fig. 4). We hope the toolkit aligns with the needs of data publishers and facilitates a smoother and more user-friendly process of data management and publishing.

© Liu M et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

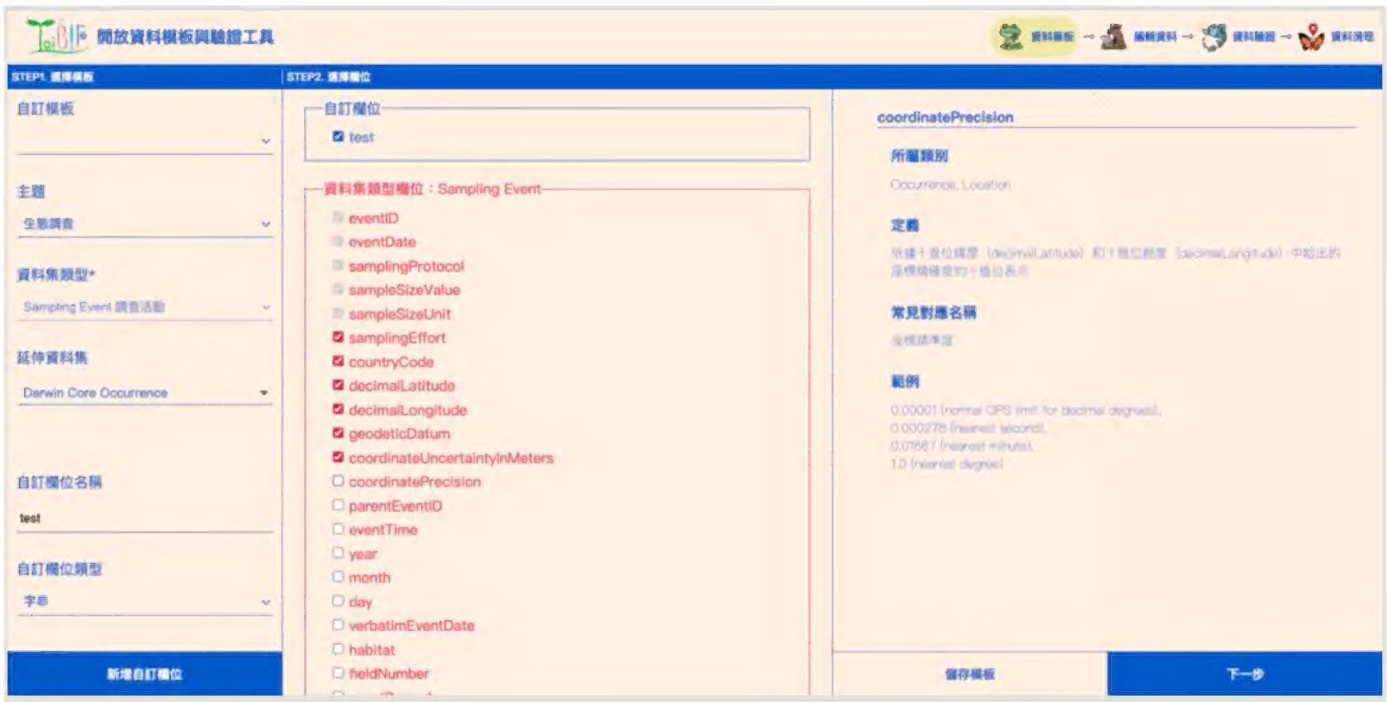

Figure 1.

Step 1 & 2: Choose a dataset type and generate the data template (screenshot from TaiBIF ODT).

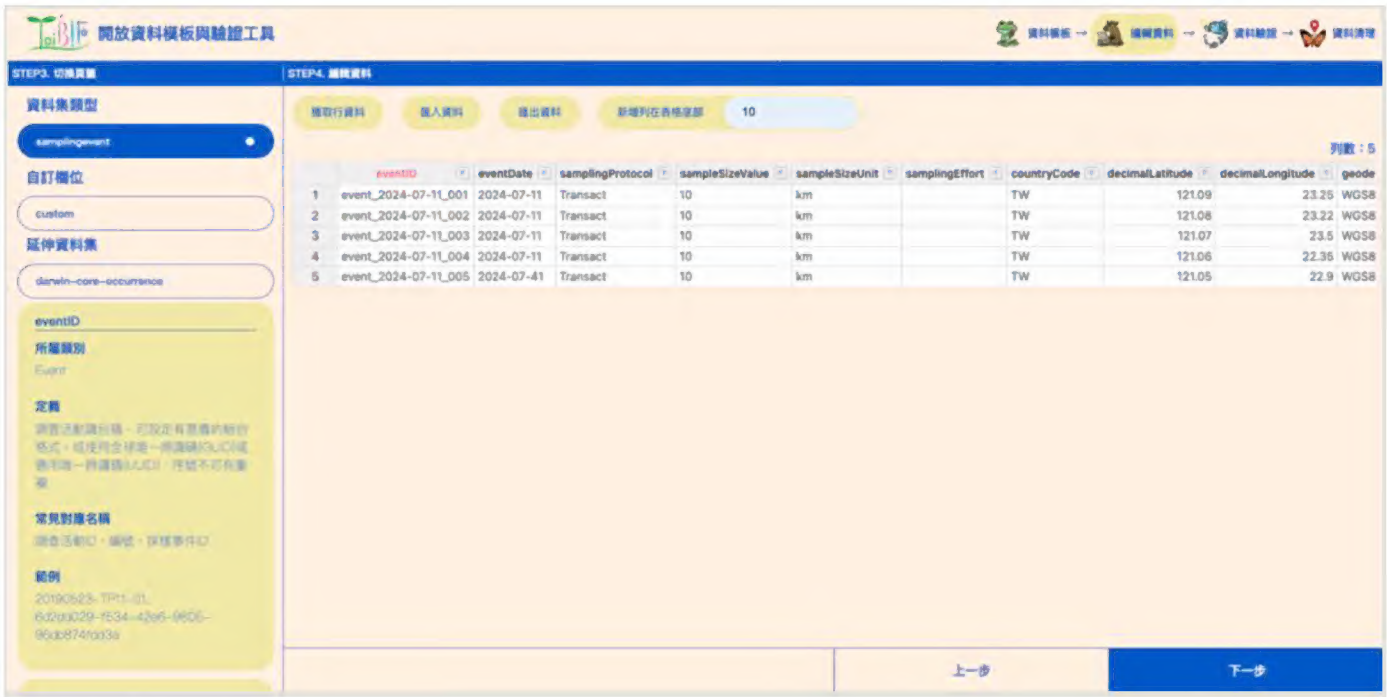

Figure 2.

Step 3 & 4: Choose core and extensions to edit data (screenshot from TaiBIF ODT).

Keywords

data template, data cleaning, data validation, biodiversity data tool

Presenting author

Jhu-Jyun Jhang

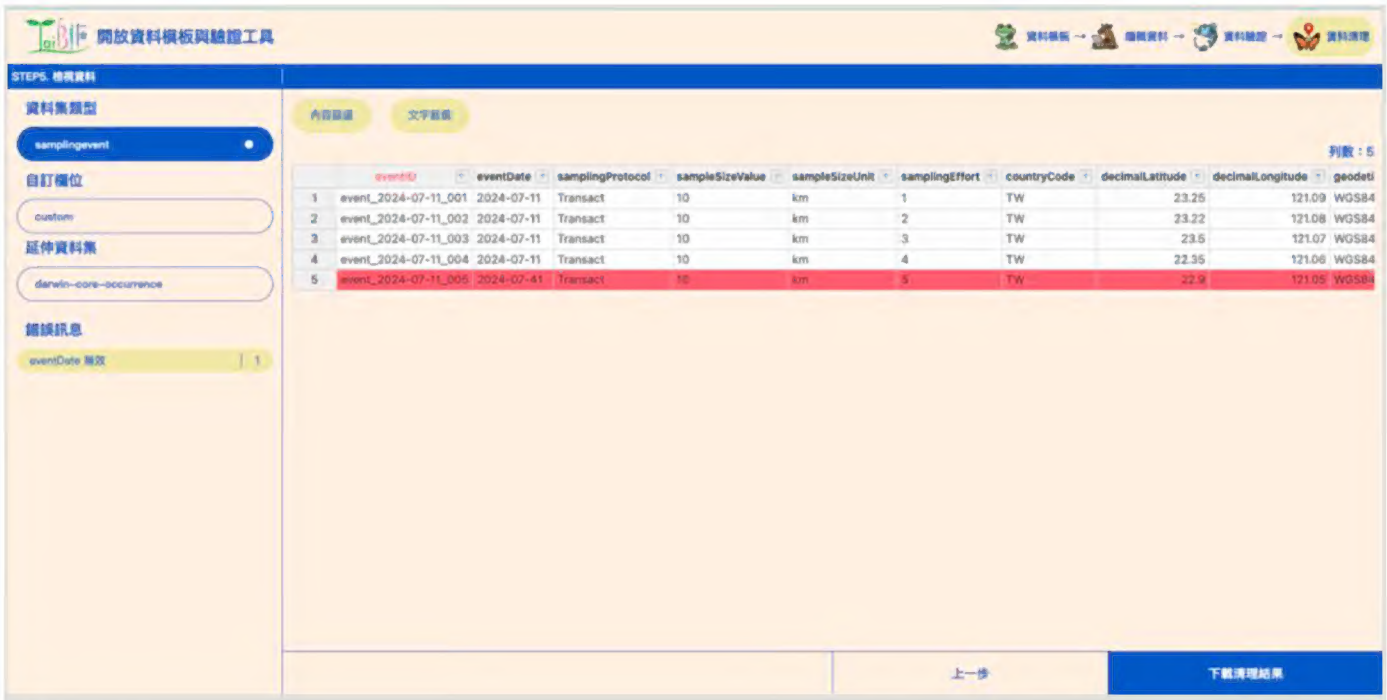

Figure 3.

Step 5: Validate the dataset. The records that have issues will be shown in red (screenshot from TaiBIF ODT).

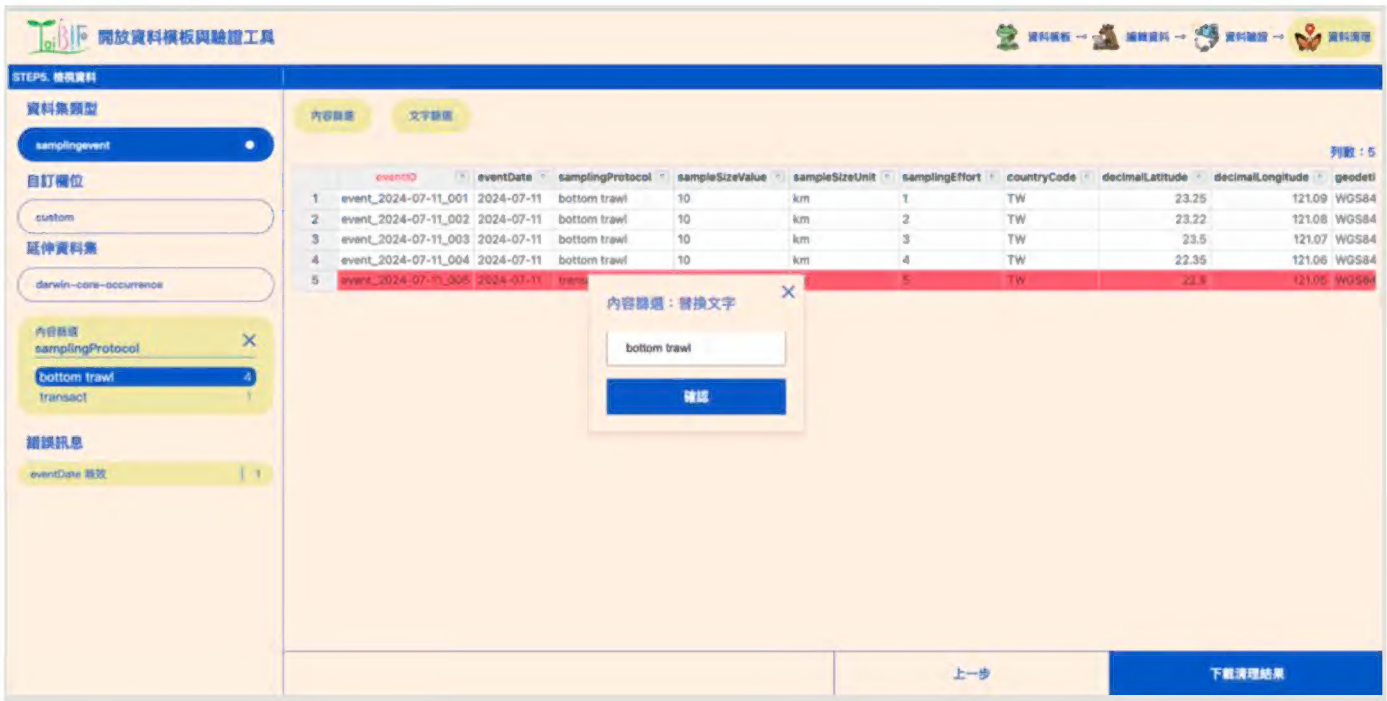

Figure 4.

Step 6: Clean and export the dataset. It provides useful features such as bulk changes, text filters, text facets, etc. (screenshot from TaiBIF ODT).

Presented at

SPNHC-TDWG 2024

Acknowledgements

This project is funded by the Taiwan Ecological Network and the Forestry and Nature Conservation Agency, Ministry of Agriculture, Taiwan.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Chapman AD (2005) Principles of Data Quality. Global Biodiversity Information Facility, Copenhagen. [In English, French, Spanish, Traditional Chiniese, Korean, Portuguese]. URL: https://www.gbif.org/document/80509/principles-of-data-quality
- Darwin Core Task Group (2009) Darwin Core. Biodiversity Information Standards (TDWG). URL: http://www.tdwg.org/standards/450